

# EDAv2 Kelompok 1 2023

October 1, 2023

Sebuah studi dilakukan untuk memahami bagaimana UMKM (Usaha Mikro, Kecil, dan Menengah) berperan dalam tingkat kemiskinan di berbagai kota dan kabupaten. Data yang digunakan untuk analisis ini mencakup informasi tentang jumlah UMKM, pendapatan UMKM, tingkat kemiskinan, dan variabel terkait lainnya. Data yang digunakan dimulai dari tahun 2019 sampai dengan 2021.

## 1 Kelompok 1

1. Meiva Labibah Putri (2204343)
2. Nabil Hanif Achmaddiredja (2205905)
3. Tattha Maharany Yasmin Akbar (2201805)
4. Muhammad Yusdan Ali Batubara (2206847)
5. Ahmad Taufiq Hidayat (2202074)

## 2 Praproses Data

Memperbaiki atribut

1. Memperbaiki atribut pada setiap kolom dari uppercase menjadi title

Mengubah datatype

1. Mengubah datatype tahun dari int menjadi datetime

Transformasi data

1. Mengubah atribut Makanan pada kolom kategori\_usaha menjadi Makanan & Minuman
2. Mengubah atribut Minuman pada kolom kategori\_usaha menjadi Makanan & Minuman
3. Mengubah atribut Batik pada kolom kategori\_usaha menjadi Fashion
4. Pembatasan rentang tahun pada dataframe Kemiskinan dari tahun 2019 - 2021

Melakukan drop pada atribut

1. Penghapusan atribut yang redundan

```
[ ]: %matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

## 2.1 Load Dataset

```
[ ]: df_umkm = pd.  
      ↳read_excel("jml_ush_mikro_kecil_menengah_umkm__kabupatenkota_kateg_data.  
      ↳xlsx")  
df_umkm.head()
```

```
[ ]:      id  kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \  
0      1           32     JAWA BARAT           3201     KABUPATEN BOGOR  
1      2           32     JAWA BARAT           3201     KABUPATEN BOGOR  
2      3           32     JAWA BARAT           3201     KABUPATEN BOGOR  
3      4           32     JAWA BARAT           3201     KABUPATEN BOGOR  
4      5           32     JAWA BARAT           3201     KABUPATEN BOGOR  
  
      kategori_usaha  jumlah_umkm  satuan  tahun  
0      AKSESORIS           927    UNIT    2017  
1      BATIK           927    UNIT    2017  
2      BORDIR          132    UNIT    2017  
3      CRAFT        33111    UNIT    2017  
4      FASHION       32316    UNIT    2017
```

```
[ ]: df_penduduk = pd.  
      ↳read_excel("jumlah_penduduk_berdasarkan_kepemilikan_ktp_elektronik_data.  
      ↳xlsx")  
df_penduduk.head()
```

```
[ ]:      id  kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \  
0      1           32     JAWA BARAT           3201     KABUPATEN BOGOR  
1      2           32     JAWA BARAT           3202     KABUPATEN SUKABUMI  
2      3           32     JAWA BARAT           3203     KABUPATEN CIANJUR  
3      4           32     JAWA BARAT           3204     KABUPATEN BANDUNG  
4      5           32     JAWA BARAT           3205     KABUPATEN GARUT  
  
      jumlah_penduduk  satuan  tahun  
0      2510633      JIWA    2015  
1      1349659      JIWA    2015  
2      1159385      JIWA    2015  
3      1765479      JIWA    2015  
4      824946      JIWA    2015
```

```
[ ]: df_pendudukmiskin = pd.  
      ↳read_excel("persentase_penduduk_miskin_berdasarkan_kabupatenkota_data.xlsx")  
df_pendudukmiskin = df_pendudukmiskin.rename(columns={'persentase_penduduk':  
      ↳'persentase_penduduk_miskin'})  
df_pendudukmiskin.head()
```

```
[ ]:   id  kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \
0    1             32    JAWA BARAT             3201    KABUPATEN BOGOR
1    2             32    JAWA BARAT             3202    KABUPATEN SUKABUMI
2    3             32    JAWA BARAT             3203    KABUPATEN CIANJUR
3    4             32    JAWA BARAT             3204    KABUPATEN BANDUNG
4    5             32    JAWA BARAT             3205    KABUPATEN GARUT

      persentase_penduduk_miskin  satuan  tahun
0                9.97  PERSEN    2010
1               10.65  PERSEN    2010
2               14.32  PERSEN    2010
3                9.30  PERSEN    2010
4               13.94  PERSEN    2010
```

## 2.2 Memperbaiki nama atribut di tiap dataframe

```
[ ]: df_umkm['nama_provinsi'] = df_umkm['nama_provinsi'].str.title()
df_umkm['nama_kabupaten_kota'] = df_umkm['nama_kabupaten_kota'].str.title()
df_umkm['kategori_usaha'] = df_umkm['kategori_usaha'].str.title()
df_umkm['satuan'] = df_umkm['satuan'].str.title()
df_umkm.head()
```

```
[ ]:   id  kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \
0    1             32    Jawa Barat             3201    Kabupaten Bogor
1    2             32    Jawa Barat             3201    Kabupaten Bogor
2    3             32    Jawa Barat             3201    Kabupaten Bogor
3    4             32    Jawa Barat             3201    Kabupaten Bogor
4    5             32    Jawa Barat             3201    Kabupaten Bogor

      kategori_usaha  jumlah_umkm  satuan  tahun
0      Aksesoris           927    Unit    2017
1        Batik           927    Unit    2017
2       Bordir           132    Unit    2017
3        Craft        33111    Unit    2017
4       Fashion        32316    Unit    2017
```

```
[ ]: df_penduduk['nama_provinsi'] = df_penduduk['nama_provinsi'].str.title()
df_penduduk['nama_kabupaten_kota'] = df_penduduk['nama_kabupaten_kota'].str.
    title()
df_penduduk['satuan'] = df_penduduk['satuan'].str.title()
df_penduduk.head()
```

```
[ ]:   id  kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \
0    1             32    Jawa Barat             3201    Kabupaten Bogor
1    2             32    Jawa Barat             3202    Kabupaten Sukabumi
2    3             32    Jawa Barat             3203    Kabupaten Cianjur
3    4             32    Jawa Barat             3204    Kabupaten Bandung
```

4	5	32	Jawa Barat	3205	Kabupaten Garut
---	---	----	------------	------	-----------------

	jumlah_penduduk	satuan	tahun
0	2510633	Jiwa	2015
1	1349659	Jiwa	2015
2	1159385	Jiwa	2015
3	1765479	Jiwa	2015
4	824946	Jiwa	2015

```
[ ]: df_pendudukmiskin['nama_provinsi'] = df_pendudukmiskin['nama_provinsi'].str.
      title()
df_pendudukmiskin['nama_kabupaten_kota'] =
      df_pendudukmiskin['nama_kabupaten_kota'].str.title()
df_pendudukmiskin['satuan'] = df_pendudukmiskin['satuan'].str.title()
df_pendudukmiskin.head()
```

	id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	\
0	1	32	Jawa Barat	3201	Kabupaten Bogor	
1	2	32	Jawa Barat	3202	Kabupaten Sukabumi	
2	3	32	Jawa Barat	3203	Kabupaten Cianjur	
3	4	32	Jawa Barat	3204	Kabupaten Bandung	
4	5	32	Jawa Barat	3205	Kabupaten Garut	

	persentase_penduduk_miskin	satuan	tahun
0	9.97	Persen	2010
1	10.65	Persen	2010
2	14.32	Persen	2010
3	9.30	Persen	2010
4	13.94	Persen	2010

## 2.3 Eksplorasi dataframe UMKM dan Kemiskinan

```
[ ]: df_umkm.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1350 entries, 0 to 1349
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    1350 non-null   int64
1   kode_provinsi         1350 non-null   int64
2   nama_provinsi         1350 non-null   object
3   kode_kabupaten_kota   1350 non-null   int64
4   nama_kabupaten_kota   1350 non-null   object
5   kategori_usaha        1350 non-null   object
6   jumlah_umkm           1350 non-null   int64
7   satuan                1350 non-null   object
8   tahun                1350 non-null   int64
```

```
dtypes: int64(5), object(4)
memory usage: 95.1+ KB
```

```
[ ]: df_penduduk.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 135 entries, 0 to 134
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    135 non-null   int64
1   kode_provinsi        135 non-null   int64
2   nama_provinsi        135 non-null   object
3   kode_kabupaten_kota  135 non-null   int64
4   nama_kabupaten_kota  135 non-null   object
5   jumlah_penduduk      135 non-null   int64
6   satuan               135 non-null   object
7   tahun               135 non-null   int64
dtypes: int64(5), object(3)
memory usage: 8.6+ KB
```

```
[ ]: df_pendudukmiskin.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 346 entries, 0 to 345
Data columns (total 8 columns):
#   Column                      Non-Null Count  Dtype
---  -
0   id                          346 non-null   int64
1   kode_provinsi              346 non-null   int64
2   nama_provinsi              346 non-null   object
3   kode_kabupaten_kota        346 non-null   int64
4   nama_kabupaten_kota        346 non-null   object
5   persentase_penduduk_miskin  346 non-null   float64
6   satuan                     346 non-null   object
7   tahun                     346 non-null   int64
dtypes: float64(1), int64(4), object(3)
memory usage: 21.8+ KB
```

## 2.4 Transformasi Data UMKM

Tim berasumsi untuk kategori UMKM Makanan & Minuman adalah makanan kemasan sedangkan untuk Kuliner itu adalah makanan yang makanan cepat saji seperti di Restoran. Selanjutnya untuk kategori Batik, tim juga berasumsi itu termasuk dari bagian kategori fashion.

```
[ ]: df_umkm.kategori_usaha = df_umkm.kategori_usaha.replace({"Makanan": "Makanan & Minuman", "Minuman": "Makanan & Minuman", "Batik": "Fashion"})
```

### 2.4.1 Penggabungan kategori usaha pada dataframe UMKM

```
[ ]: df_umkm = df_umkm.groupby(['nama_provinsi', 'kode_kabupaten_kota',  
    ↳ 'nama_kabupaten_kota', 'kategori_usaha', 'satuan', 'tahun'])['jumlah_umkm'].  
    ↳ sum().reset_index()  
df_umkm = df_umkm[['nama_provinsi', 'kode_kabupaten_kota',  
    ↳ 'nama_kabupaten_kota', 'kategori_usaha', 'jumlah_umkm', 'satuan', 'tahun']]  
df_umkm.head()
```

```
[ ]:  nama_provinsi  kode_kabupaten_kota  nama_kabupaten_kota  kategori_usaha  \  
0      Jawa Barat           3201      Kabupaten Bogor      Aksesoris  
1      Jawa Barat           3201      Kabupaten Bogor      Aksesoris  
2      Jawa Barat           3201      Kabupaten Bogor      Aksesoris  
3      Jawa Barat           3201      Kabupaten Bogor      Aksesoris  
4      Jawa Barat           3201      Kabupaten Bogor      Aksesoris  
  
    jumlah_umkm  satuan  tahun  
0           927    Unit    2017  
1           984    Unit    2018  
2          1045    Unit    2019  
3          1110    Unit    2020  
4          1179    Unit    2021
```

### 2.4.2 Pembatasan DataFrame dari tahun 2019 sampai 2021

Tim melakukan pembatasan untuk bisa menyatukan data

```
[ ]: df_umkm = df_umkm[(df_umkm['tahun'] > 2018) & (df_umkm['tahun'] < 2022)]  
df_umkm = df_umkm[['nama_provinsi', 'kode_kabupaten_kota',  
    ↳ 'nama_kabupaten_kota', 'kategori_usaha', 'jumlah_umkm', 'satuan', 'tahun']]  
df_umkm.head()
```

```
[ ]:  nama_provinsi  kode_kabupaten_kota  nama_kabupaten_kota  kategori_usaha  \  
2      Jawa Barat           3201      Kabupaten Bogor      Aksesoris  
3      Jawa Barat           3201      Kabupaten Bogor      Aksesoris  
4      Jawa Barat           3201      Kabupaten Bogor      Aksesoris  
7      Jawa Barat           3201      Kabupaten Bogor      Bordir  
8      Jawa Barat           3201      Kabupaten Bogor      Bordir  
  
    jumlah_umkm  satuan  tahun  
2           1045    Unit    2019  
3           1110    Unit    2020  
4           1179    Unit    2021  
7            149    Unit    2019  
8            159    Unit    2020
```

```
[ ]:
```

```
df_penduduk = df_penduduk[(df_penduduk['tahun'] > 2018) & (df_penduduk['tahun'] < 2022)]
df_penduduk = df_penduduk[['kode_provinsi', 'nama_provinsi',
                             'kode_kabupaten_kota', 'nama_kabupaten_kota', 'jumlah_penduduk', 'satuan',
                             'tahun']]
df_penduduk.head()
```

```
[ ]:      kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \
27          32      Jawa Barat          3201      Kabupaten Bogor
28          32      Jawa Barat          3202      Kabupaten Sukabumi
29          32      Jawa Barat          3203      Kabupaten Cianjur
30          32      Jawa Barat          3204      Kabupaten Bandung
31          32      Jawa Barat          3205      Kabupaten Garut

      jumlah_penduduk satuan  tahun
27          3424309      Jiwa  2019
28          1798217      Jiwa  2019
29          1600728      Jiwa  2019
30          2413818      Jiwa  2019
31          1495395      Jiwa  2019
```

```
[ ]: df_pendudukmiskin = df_pendudukmiskin[(df_pendudukmiskin['tahun'] > 2018) &
                                             (df_pendudukmiskin['tahun'] < 2022)]
df_pendudukmiskin = df_pendudukmiskin[['kode_provinsi', 'nama_provinsi',
                                         'kode_kabupaten_kota', 'nama_kabupaten_kota', 'persentase_penduduk_miskin',
                                         'satuan', 'tahun']]
df_pendudukmiskin.head()
```

```
[ ]:      kode_provinsi nama_provinsi  kode_kabupaten_kota nama_kabupaten_kota \
238          32      Jawa Barat          3201      Kabupaten Bogor
239          32      Jawa Barat          3202      Kabupaten Sukabumi
240          32      Jawa Barat          3203      Kabupaten Cianjur
241          32      Jawa Barat          3204      Kabupaten Bandung
242          32      Jawa Barat          3205      Kabupaten Garut

      persentase_penduduk_miskin satuan  tahun
238                6.66  Persen  2019
239                6.22  Persen  2019
240                9.15  Persen  2019
241                5.94  Persen  2019
242                8.98  Persen  2019
```

## 2.5 Merge Ketiga DataFrame

```
[ ]: df2 = pd.merge(df_umkm, df_penduduk, how="left", on=["kode_kabupaten_kota", "tahun"])
df2 = pd.merge(df2, df_pendudukmiskin, how="left", on=["kode_kabupaten_kota", "tahun"])
df2.head()
```

```
[ ]:  nama_provinsi_x  kode_kabupaten_kota  nama_kabupaten_kota_x  kategori_usaha \
0      Jawa Barat      3201      Kabupaten Bogor      Aksesoris
1      Jawa Barat      3201      Kabupaten Bogor      Aksesoris
2      Jawa Barat      3201      Kabupaten Bogor      Aksesoris
3      Jawa Barat      3201      Kabupaten Bogor      Bordir
4      Jawa Barat      3201      Kabupaten Bogor      Bordir
```

```
    jumlah_umkm  satuan_x  tahun  kode_provinsi_x  nama_provinsi_y \
0          1045    Unit    2019           32      Jawa Barat
1          1110    Unit    2020           32      Jawa Barat
2          1179    Unit    2021           32      Jawa Barat
3           149    Unit    2019           32      Jawa Barat
4           159    Unit    2020           32      Jawa Barat
```

```
    nama_kabupaten_kota_y  jumlah_penduduk  satuan_y  kode_provinsi_y \
0      Kabupaten Bogor      3424309      Jiwa           32
1      Kabupaten Bogor      3579497      Jiwa           32
2      Kabupaten Bogor      3694126      Jiwa           32
3      Kabupaten Bogor      3424309      Jiwa           32
4      Kabupaten Bogor      3579497      Jiwa           32
```

```
    nama_provinsi  nama_kabupaten_kota  persentase_penduduk_miskin  satuan
0      Jawa Barat      Kabupaten Bogor              6.66  Persen
1      Jawa Barat      Kabupaten Bogor              7.69  Persen
2      Jawa Barat      Kabupaten Bogor              8.13  Persen
3      Jawa Barat      Kabupaten Bogor              6.66  Persen
4      Jawa Barat      Kabupaten Bogor              7.69  Persen
```

### 2.5.1 Penghapusan nama atribut yang terduplikasi

```
[ ]: df2 = df2.drop(columns=['nama_provinsi_x', 'nama_kabupaten_kota_x', 'satuan_x',
    ↪ 'nama_provinsi_y', 'nama_kabupaten_kota_y', 'satuan_y', 'kode_provinsi_y',
    ↪ 'satuan'])
df2 = df2.rename(columns={'kode_provinsi_x': 'kode_provinsi'})
df2.head()
```

```
[ ]:  kode_kabupaten_kota  kategori_usaha  jumlah_umkm  tahun  kode_provinsi \
0          3201      Aksesoris          1045    2019           32
1          3201      Aksesoris          1110    2020           32
```



2	3201	Aksesoris	1179	2021	32
3	3201	Bordir	149	2019	32
4	3201	Bordir	159	2020	32

	jumlah_penduduk	nama_provinsi	nama_kabupaten_kota	\
0	3424309	Jawa Barat	Kabupaten Bogor	
1	3579497	Jawa Barat	Kabupaten Bogor	
2	3694126	Jawa Barat	Kabupaten Bogor	
3	3424309	Jawa Barat	Kabupaten Bogor	
4	3579497	Jawa Barat	Kabupaten Bogor	

	persentase_penduduk_miskin
0	6.66
1	7.69
2	8.13
3	6.66
4	7.69

```
[ ]: df2 = df2[['kode_provinsi', 'nama_provinsi', 'kode_kabupaten_kota',
↳ 'nama_kabupaten_kota', 'kategori_usaha', 'jumlah_umkm', 'jumlah_penduduk',
↳ 'persentase_penduduk_miskin', 'tahun']]
df2.head()
```

	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	\
0	32	Jawa Barat	3201	Kabupaten Bogor	
1	32	Jawa Barat	3201	Kabupaten Bogor	
2	32	Jawa Barat	3201	Kabupaten Bogor	
3	32	Jawa Barat	3201	Kabupaten Bogor	
4	32	Jawa Barat	3201	Kabupaten Bogor	

	kategori_usaha	jumlah_umkm	jumlah_penduduk	persentase_penduduk_miskin	\
0	Aksesoris	1045	3424309	6.66	
1	Aksesoris	1110	3579497	7.69	
2	Aksesoris	1179	3694126	8.13	
3	Bordir	149	3424309	6.66	
4	Bordir	159	3579497	7.69	

	tahun
0	2019
1	2020
2	2021
3	2019
4	2020

```
[ ]:
```

```
df3 = df2.groupby(['kode_kabupaten_kota', 'nama_kabupaten_kota',
    ↳ 'jumlah_penduduk', 'persentase_penduduk_miskin', 'tahun'])['jumlah_umkm'].
    ↳ sum().reset_index()
df3.head()
```

```
[ ]:  kode_kabupaten_kota  nama_kabupaten_kota  jumlah_penduduk \
0          3201      Kabupaten Bogor          3424309
1          3201      Kabupaten Bogor          3579497
2          3201      Kabupaten Bogor          3694126
3          3202  Kabupaten Sukabumi          1798217
4          3202  Kabupaten Sukabumi          1815278

    persentase_penduduk_miskin  tahun  jumlah_umkm
0                6.66    2019          449058
1                7.69    2020          476843
2                8.13    2021          506347
3                6.22    2019          322086
4                7.09    2020          342014
```

### 3 Visualisasi

Tim menggunakan visualisasi data untuk membantu memahami pola dan hubungan antara UMKM dan tingkat kemiskinan. Membuat grafik sebaran yang menunjukkan hubungan antara persentase UMKM dan tingkat kemiskinan, serta grafik batang yang menggambarkan jumlah UMKM dengan jumlah penduduk yang mempunyai KTP elektronik di berbagai wilayah.

```
[ ]: df4 = df2.groupby(['nama_kabupaten_kota', 'jumlah_penduduk',
    ↳ 'tahun'])['jumlah_umkm'].sum().reset_index()
df5 = df2.groupby(['nama_kabupaten_kota', 'kategori_usaha',
    ↳ 'tahun'])['jumlah_umkm'].sum().reset_index()
df6 = df2.groupby(['nama_kabupaten_kota', 'jumlah_penduduk',
    ↳ 'persentase_penduduk_miskin', 'tahun'])['jumlah_umkm'].sum().reset_index()
df5_2019 = df5[df5['tahun'] == 2019]
df5_2020 = df5[df5['tahun'] == 2020]
df5_2021 = df5[df5['tahun'] == 2021]
df6_2021 = df6[df6['tahun'] == 2021]
```

```
[ ]: df_filter = df3[['nama_kabupaten_kota', 'jumlah_penduduk', 'tahun']]

# Pivot the DataFrame to have years as columns
df_pivot = df_filter.pivot(index='nama_kabupaten_kota', columns='tahun',
    ↳ values='jumlah_penduduk')

# Plotting
fig, ax = plt.subplots(figsize=(15, 10))
```

```

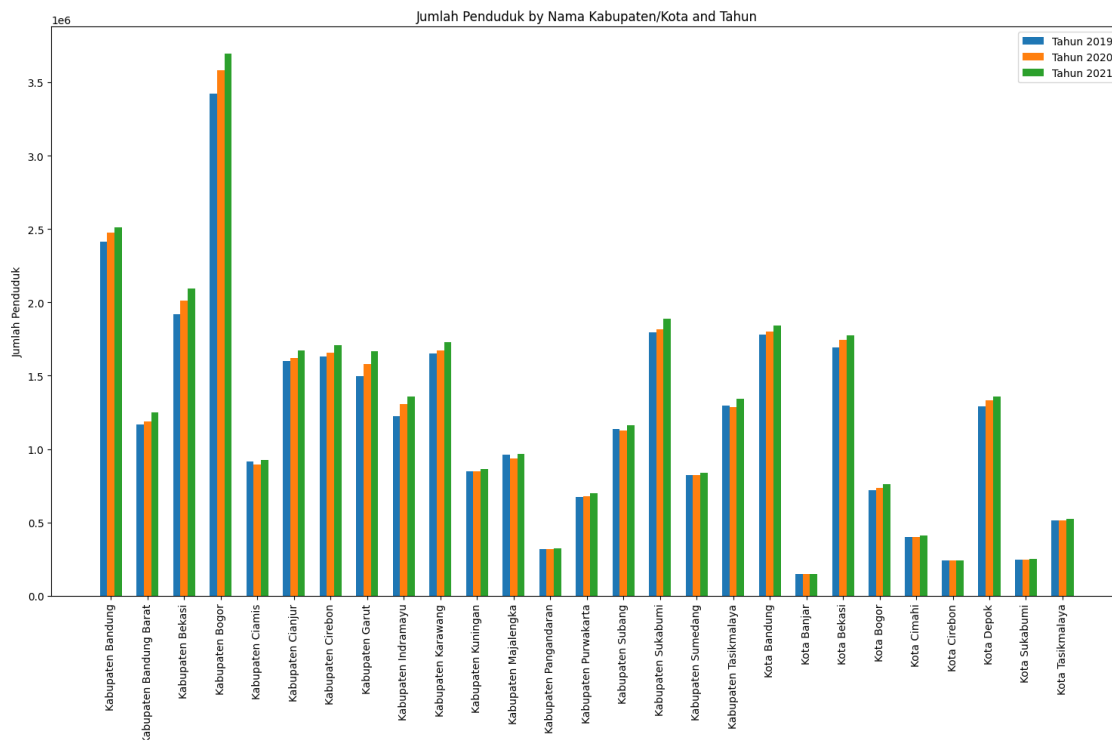
years = df_pivot.columns.tolist()
x = range(len(df_pivot.index))
width = 0.2

for i, year in enumerate(years):
    ax.bar(
        [pos + i * width for pos in x],
        df_pivot[year],
        width=width,
        label=f'Tahun {year}'
    )

ax.set_ylabel('Jumlah Penduduk')
ax.set_title('Jumlah Penduduk by Nama Kabupaten/Kota and Tahun')
ax.set_xticks([pos + width for pos in x])
ax.set_xticklabels(df_pivot.index, rotation=90)
ax.legend()

plt.tight_layout()
plt.show()

```



```
[ ]: df_filter2 = df3[['nama_kabupaten_kota', 'jumlah_umkm', 'tahun']]
```

```

# Pivot the DataFrame to have years as columns
df_pivot = df_filter2.pivot(index='nama_kabupaten_kota', columns='tahun',
    ↪values='jumlah_umkm')

# Plotting
fig, ax = plt.subplots(figsize=(15, 10))

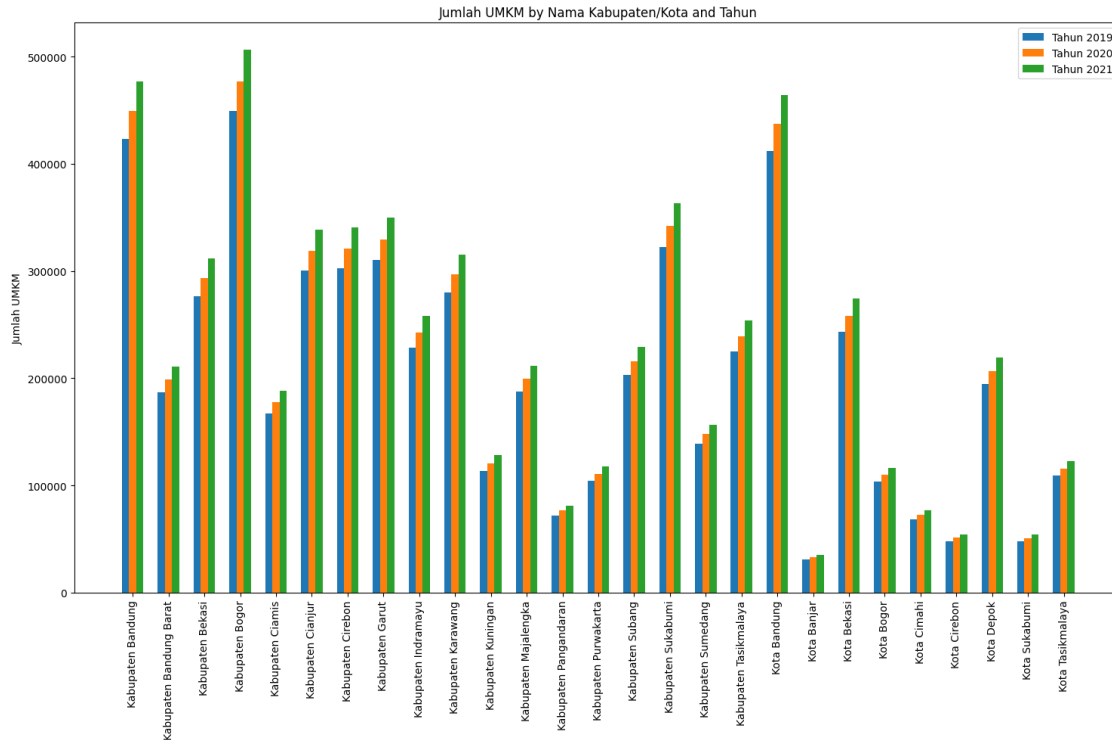
years = df_pivot.columns.tolist()
x = range(len(df_pivot.index))
width = 0.2

for i, year in enumerate(years):
    ax.bar(
        [pos + i * width for pos in x],
        df_pivot[year],
        width=width,
        label=f'Tahun {year}'
    )

ax.set_ylabel('Jumlah UMKM')
ax.set_title('Jumlah UMKM by Nama Kabupaten/Kota and Tahun')
ax.set_xticks([pos + width for pos in x])
ax.set_xticklabels(df_pivot.index, rotation=90)
ax.legend()

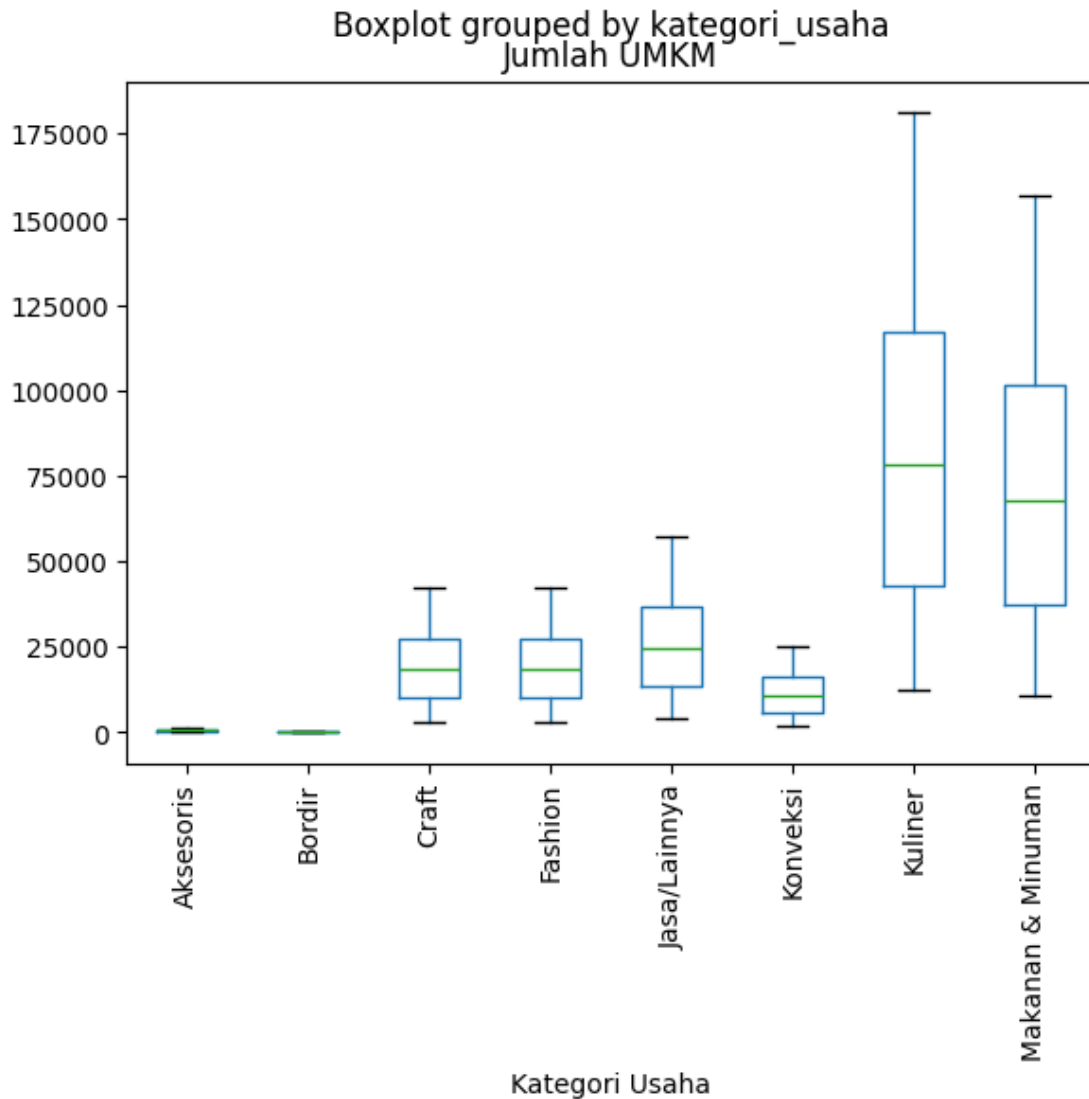
plt.tight_layout()
plt.show()

```



Boxplot untuk menampilkan atribut kategori usaha berdasarkan jumlah umkm, disini berfokus pada jumlah UMKM di tahun 2021 karena data pada tahun ini adalah data yang paling terbaru

```
[ ]: df5_2021.boxplot(by = 'kategori_usaha', column = ['jumlah_umkm'], grid = False)
plt.title("Jumlah UMKM")
plt.xlabel('Kategori Usaha')
plt.xticks(rotation=90)
plt.show()
```



```
[ ]: import warnings
warnings.filterwarnings('ignore')

plt.figure(figsize=(12, 8))

# Scatter plot
sns.scatterplot(data=df6_2021, x='persentase_penduduk_miskin', y='jumlah_umkm',
               size='jumlah_penduduk', hue='jumlah_penduduk', sizes=(20, 400),
               legend='full')

# Annotate city names
for i, row in df6_2021.iterrows():
```

```

plt.annotate(row['nama_kabupaten_kota'],
            ↪(row['persentase_penduduk_miskin'], row['jumlah_umkm']))

plt.title('Scatter plot of Persentase Penduduk Miskin vs Jumlah UMKM vs Jumlah_
            ↪Penduduk by Kabupaten/Kota')
plt.xlabel('Persentase Penduduk Miskin')
plt.ylabel('Jumlah UMKM')
plt.show()

```

